Clinical Radiology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Review

Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas

P. Korfiatis, B. Erickson*

Department of Radiology, Mayo Clinic, 200 1st Street SW, Rochester, MN, 55905, USA

This paper describes state-of-the-art methods for molecular biomarker prediction utilising magnetic resonance imaging. This review paper covers both classical machine learning approaches and deep learning approaches to identifying the predictive features and to perform the actual prediction. In particular, there have been substantial advances in recent years in predicting molecular markers for diffuse gliomas. There are few examples of molecular marker prediction for other brain tumours. Deep learning has contributed significantly to these advances, but suffers from challenges in identifying the features used to make predictions. Tools to better identify and understand those features represent an important area of active research.

Introduction

There has been a revolution in the world of computer science with new formulations of deep learning technology achieving performance that exceeds humans in the identification of content in images. This revolution has caused some to predict that computers might replace radiologists.[1] Although that has generated many vigorous discussions, these new formulations have the ability to determine decision criteria using only the "raw" image without human-informed selection and computation of the features in the image that are thought to be important. This means that computers may identify aspects of images that are unperceivable by humans or are too subtle to have been recognised. In this article we discuss the use of new machine learning methods (deep learning) to predict molecular properties of tissues using routine radiological images, a task that is usually outside the scope of normal radiological practice.

Machine learning and deep learning

Supervised machine learning is a family of techniques that take many examples with known properties (e.g., they do or do not have disease X) and uses properties (features) extracted from those examples to "learn" the correct combination of property weighting to most accurately predict the "class". Supervised machine learning methods have several important factors in common: they require "features" that are the fundamental elements of an image which must be learned in order to perform the task. Although these were identified and computed by the scientist when using traditional machine learning tools, the computer learns the features in most deep learning applications. This means that features that are not intuitive or not known, might be discovered.

Convolutional neural network (CNNs) are especially useful for medical image analysis as the core idea is to use convolutions to generate features that capture the key features of the images. An important benefit of CNN deep learning algorithms, as compared with traditional machine learning methods, is that there is no need to compute

* Guarantor and correspondent: B. Erickson, Mayo Clinic, 200 1st St SW, Rochester, MN 55905, USA. Tel.: (507) 774-6238.
E-mail address: bje@mayo.edu (B. Erickson).

https://doi.org/10.1016/j.crad.2019.01.028
0009-9260/

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028

features as a first step. The CNN effectively finds the important features as part of its search process, eliminating the steps of feature engineering and selection, which are parts of classical machine learning. Fig 1 captures a version of a simple CNN architecture.

All supervised machine learning methods face the challenge of needing enough examples to assure the algorithm learns the general principles of the images, and not the specific example images given. This challenge of "generalisation" is greater for deep learning because it has a much larger parameter space, which allows for more ability to learn the specific examples. Therefore it is much more important to use methods that assure generalisation of the machine learning system.

Radiogenomics

Radiogenomics has been used to mean two different things: the earliest use of this term described the sensitivity or effectiveness of radiation therapy and the impact of genomic properties of the tumour being treated. The other definition, which is the one we will focus on in this article, is the prediction of genomic properties of tissues using radiological imaging. The traditional role of radiological imaging has been to image the phenotype of a patient—that is, to quantify and assess the product of the genotype of a subject, as well as the impact of environment (e.g., trauma) or habits (exercise and eating) on that individual. We note here that we will use the term "molecular" rather than "genomic" because many important findings are technically not genomic, but may be epigenomic (e.g., methylation of a portion of the genome, which then typically deactivates that gene) or chromosomal or molecular level, but not actually representing a change in the genetic make-up.

There have been publications recognising tendencies between imaging findings and molecular properties. Some examples include the recognition that 1p/19q chromosomal co-deletion in oligodendroglioma often resulted in ill-defined margins and heterogeneous T2 signal. Although these can produce statistically significant *p*-values,[2] prospective prediction of genomic properties in a specific case was not accurate enough for clinical utility.

We also note here that most of the techniques used for radiogenomic predictions can be applied to other imaging tasks, including the classification of diseases (benign, malignant, inflammatory, infectious, etc.) or the assessment of treatment response. The common point is that the algorithms are able to learn patterns in the images that correlate with the class.

Current state of the art for radiogenomic predictions for brain tumours

Table 1 summarises state-of-the-art publications focusing on radiogenomics and more specifically using machine learning to predict molecular biomarkers utilising magnetic resonance imaging (MRI). The majority of the

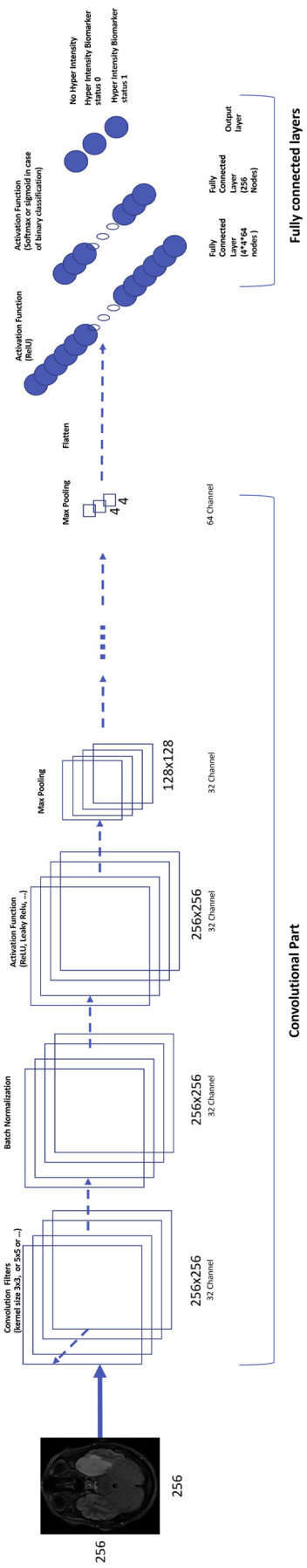

Figure 1 Example of a CNN architecture for multiclass classification utilising fluid attenuation inversion recovery (FLAIR) MRI images as input. The network consists of two parts: one part for generating features (convolutional part) and a part dedicated to performing the classification based on the input from the above layers. The convolution parts consist of convolution layers, batch normalisation, activation, and maxpooling layers, components commonly utilised in CNN architectures.

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028

Table 1
Recent publications that describe machine learning for predicting molecular biomarkers utilising MRI in brain tumours.

Reference	Marker	Dataset	Tumour	Method	Results
Jakola[15]	IDH	25, FLAIR	Grade 2 & 3 diffuse glioma	Texture analysis, grey level co occurrence	Homogeneity, AUC 0.905
Li[16]	MGMT	193, T1 FLAIR	Diffuse glioma	Grey-level co-occurrence matrix, grey-level run length matrix, grey-level size zone matrix and neighbourhood grey-tone difference matrix methods (pyradiomics)	Six selected features, Random Forest, AUCI 0.88
Park[17]	IDH, 1p/19q	93, T2, ADC, FA	Grade 2 glioma	First order, grey level co-occurrence, least absolute shrinkage and selection operator	AUC 0.807
Chang[18]	IDH	496, FLAIR, T2, T1 (pre and post)	Diffuse glioma	Deep neural network (ResNet, 34 layers, stepped decay), age	AUC 0.95
Rathore[19]	IDH	261, FLAIR, T1 CE	Diffuse glioma	Texture features, T1, T2, Flair, DTI, perfusion, k-means clustering	The subtypes provided risk-stratification substantially beyond that provided by WHO classifications.
R. van der Voort[20]	1p19q	63,T1, T2W	Grade 2 & 3 Diffuse glioma	SVM, shape features, Gabor features,	Accuracy 95th percentile confidence interval [0.56; 0.80]
Han[21]	MGMT	26, T1, T2, FLAIR	Glioblastoma	Convolution recurrent neural network	AUC 0.61 (or 0.73 for MRI)
Korfiatis[22]	MGMT	155, T2	Glioblastoma	ResNet50	Accuracy 94.9±3.92%
Korfiatis[23]	ATRX	135 FLAIR	Glioblastoma	Custom ResNet, SVM second order statistics	ResNet f1 score 0.91 SVM f1 score 0.73
Chang[24]	IDH, 1p19q, MGMT	259, T1, FLAIR	Diffuse glioma	Custom ResNet	Accuracy 94% IDH, 92% 1p19q, 83 % MGMT
Eichinger[25]	IDH	79	Grade 2 & 3 diffuse glioma	Local binary pattern, size	AUC 0.952
Bisdas[26]	IDH	37	Grade 2 & 3 Diffuse glioma	Texture analysis (first-order statistics), SVM	AUC 0.88
Zhang[14]	IDH, TP53	103	Grade 2 & 3 diffuse glioma	SVM, 12 histogram-based textures, 9 grey-Level co-occurrence matrix features, 13 grey-level gradient co-occurrence matrix features, 13 grey-level run-length matrix features, 13 grey-level size zone matrix features, and five neighbourhood grey-tone difference matrix features	IDH AUC 0.792. TP53 AUC 0.869
Ren[27]	IDH, ATRX	57	Diffuse glioma	SVM, Second order statistics, T2 FLAIR, CBF, ADC, and eADC	IDH AUC 0.93
Pan[28]	H3K27M	151	Diffuse glioma	Grey-level co-occurrence matrix, local binary patterns, histogram of oriented gradients, and Haralick textural features	AUC 0.829
Arita[29]	IDH/TERT	169	Grade 2 & 3 diffuse glioma	Radiomics features lasso regression, T1-, T2-weighted, FLAIR, and gadolinium-enhanced T1-weighted images	IDH, accuracy 0.82, Subtype prediction 0.56
Liu[30]	H3 K27M mutation status	55	Brainstem gliomas	CNN features and support-vector-machine classifier	Accuracy 94.85%
Zhou[31]	SHH, wnt, group 3 & 4	66	Medulloblastoma	590 textures, SVM	AUC 0.7–0.8

IDH, isocitrate dehydrogenase mutations; FLAIR, fluid attenuated inversion recovery; AUC, area under the receiver operating characteristic curve; MGMT, methyltransferase promoter methylation; ADC, apparent diffusion coefficient; FA, Fractional anisotropy; CE, contrast enhanced; DTI, diffusion tensor imaging; WHO, World Health Organization; SVM, Support vector machine; MRI, magnetic resonance imaging; TP53, tumor protein p53 gene; ATRX, Alpha thalassemia/mental retardation syndrome X-linked; CBF, cerebral blood flow; eADC, exponential Apparent diffusion coefficient; TERT, Telomerase reverse transcriptase; CNN, convolutional neural network; SHH, sonic hedgehog.

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028

studies are focusing on diffuse gliomas and their isocitrate dehydrogenase (IDH), methylguanine-DNA methyltransferase (MGMT), telomerase reverse transcriptase (TERT) and chromosome arms 1p and 19q (1p19q) status, though EGFR and TP53 have also been reported. Very few studies of other types of brain tumours exist. The interest in molecular markers has been spurred by the World Health Organization (WHO) changing its recommendation for the way diffuse gliomas should be reported.[3] Specifically, the WHO recommends that diffuse gliomas have IDH mutation status, 1p19q chromosomal deletion, and TERT amplification be reported and histological properties are secondary. An article in the *New England Journal of Medicine*[4] demonstrated the value of using these molecular biomarkers to group gliomas that have similar clinical behaviours, therapy responses, and outcomes. Identification of these biomarkers may enable patient-specific treatment, resulting in improved outcomes.

Methylation of MGMT has also been shown to be an important prognostic marker in gliomas. MGMT is a DNA repair enzyme and as noted before, methylation deactivates a gene. As some types of therapies (e.g., radiation and alkylating agents) work by disrupting the DNA of a tumour, these therapies will be more effective in tumours that have this DNA repair enzyme deactivated. Thus, MGMT methylation generally has a positive survival effect.[5,6]

There is great excitement that applying modern machine learning methods to medical images may substantially aid in improving medical diagnosis. The ability to determine brain tumour genomics from routine imaging would be helpful in cases where testing is inaccurate (5–10% of 1p19q Fluorescence in situ hybridization (FISH) tests are false positives) or fails to provide a result due to inadequate sample. Tumours are genetically heterogeneous, and imaging may also provide a better "global assessment" than tissue biopsy as well as potentially helping to characterise the host response to the tumour. An imaging biomarker may help to further categorise subjects, potentially serving as a useful stratifier for clinical trials. Imaging can supplement or confirm histopathology for initial diagnosis and treatment, and may also provide insight into the impact of therapy on tumours. Some believe that simple volumetrics may indicate when Low grade glioma (LGG) are likely to convert to glioblastoma multiforme (GBM)[7] but this is controversial,[8] and more sophisticated quantitative/machine learning methods may improve our understanding of this challenge. In sum, there is likely substantial information available in routine clinical imaging that can be extracted using machine learning methods.

Connections between imaging phenotype and tumour molecular markers have been described by us and others for diffuse gliomas.[9,10] MRI features have been found to predict survival and molecular data in lower-grade gliomas.[11] Classical machine learning has utilised imaging features to predict IDH and TP53 mutation status from multimodal MRI features with moderate accuracy.[12,13] Recently, studies utilising deep learning aimed at classification and segmentation have demonstrated success in mammographic tumour classification, interstitial lung disease classification, brain cancer classification, skin cancer classification, and retinopathy image segmentation. Deep learning has become the technology of choice for numerous image-based competitions both inside medicine, and more broadly in computer science, resulting in substantial jumps in performance on these various challenges.

The methodologies summarised in Table 1 fall into two categories: classical machine learning and deep learning. A classical machine learning workflow includes feature calculation and selection performed by a human based on their insight into the disease and imaging. Then a classifier is used to predict the molecular status. Fig 2 captures the classical machine learning approach—a series of steps often referred to as a "pipeline". The majority of classical machine learning approaches require a region of interest (ROI) to be placed around the glioma in order to calculate the features. The features calculated using this type of approach can be categorised into the following three categories: (1) first order (intensity and shape based features), (2) second order (texture-based features such as grey level co-occurrence

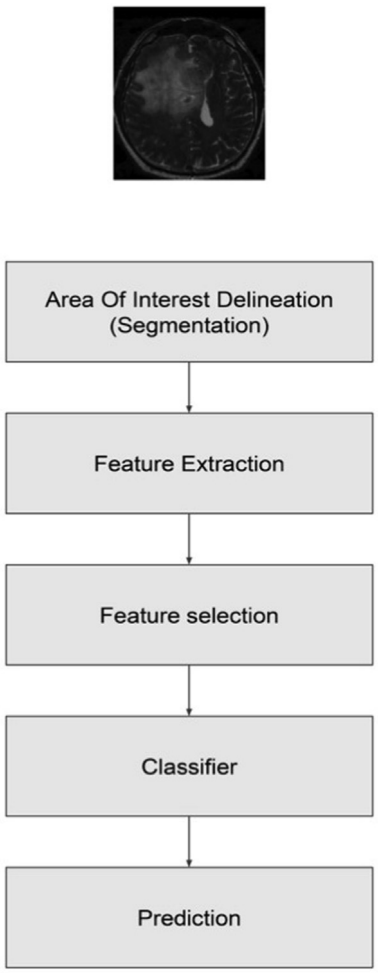

Figure 2 The steps in applying radiogenomic analysis to images. This includes identification of the tumour and possibly other regions of interest (e.g., normal tissue to serve as a reference), feature extraction, feature selection, classification, and then making the prediction based on the classifier.

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028

and grey level run-length), (3) higher-order features (filters are used to extract the features, e.g., wavelet, fractal and Laplacian of Gaussian). After calculating the features, feature selection and classification techniques are applied.

Support vector machine (SVM) is the most common classifier choice in classical machine learning approaches. Random forest classifier (RFC) is another popular choice, as it results in a more human-friendly decision tree that can provide an understanding of how the system makes a choice. The naive Bayes classifier is also a popular method, and relies on probability calculations for making decisions.

Chang *et al.*[22] reported accuracies of 94%, 92%, 83% for IDH, 1p19q and MGMT, respectively, using a dataset consisting of 259 cases. In another study utilising a residual convolution network on the largest dataset reported to date (496 patients), Chang *et al.*[16] achieved an area under the receiver operating curve (AUC) of 0.95 for IDH prediction with an accuracy of 0.891.

The majority of methods reported require manual tumour delineation, in addition to preprocessing steps such as intensity normalisation or N4 bias corrections. Korfiatis *et al.*[20] achieved high accuracy in a study aiming to predict MGMT methylation without the requirement of tumour segmentation.

There are reports of visually accessible features of non-glioma brain tumours that predict molecular subtypes for brain tumours other than diffuse glioma. Paediatric brain tumours are much less common than adult diffuse gliomas, and thus it is more challenging to get enough examples to properly train a machine learning system. There are reports using radiomics to differentiate the three common paediatric brain tumours (medulloblastoma, ependymoma, brain stem glioma) with moderate performance (AUC ranging from 0.7–0.9 depending on the study and the brain tumour type).[30,31] There are four genetically defined subtypes of medulloblastoma: sonic hedgehog (SHH)-activated (and these may or may not have TP53 mutation), wingless (WNT)-activated, and there are two non-SHH/non-WNT groups simply referred to as Groups 3 and 4. Each of these types have imaging features that can help to predict the genomic make-up.[32,33] An abstract[29] describing the use of machine learning to predict these genetic subtypes has been presented as well, but we are not aware of a peer-reviewed publication on this topic.

Although meningiomas are quite common, the application of radiomics to their images is rather uncommon. There are reports using radiomic methods to predict grade[34] and stiffness,[35] but as with paediatric tumours, we are not aware of one focused on the prediction of molecular properties.

Similar approaches have been utilised to predict tumour grade from MRI as well as differentiation between primary central nervous system lymphoma and glioblastoma.[36] Zacharachi *et al.*[37] utilised texture features and supervised classifier to different types of brain tumours and for glioma grading. Citak-Er *et al.*[38] implemented a multimodal approach and a SVM classifier achieved accuracy of 0.93.

Challenges to adoption

The most recent works in this field focus on the application of deep neural networks with several works reporting accuracies higher than the traditional machine learning approach. For the majority of works, identification of the region of interest is required before classification is performed, but if that must be performed by a human, it will increase variability and impede clinical adoption; however, approaches that utilise the entire image without a discrete step to localise tumour have been described,[21] and this may also benefit if important information is outside the area of apparent tumour.

The limited number of brain tumour datasets with known molecular markers has impeded research. This problem is amplified if there is a significant imbalance in the frequency of those markers. We noted before that a typical CNN architecture consists of a large number of parameters. This demands a larger dataset than traditional machine learning to assure that the system is generalising, and not learning the specific examples used in training. Two common techniques are utilised in order to address this issue are transfer learning and data augmentation.

Transfer learning is a useful training tool for CNN's in cases where the number of training examples is limited. Networks that have been trained for one task (e.g., recognising tomatoes in an image) can be modified to recognise similar objects (e.g., watermelons) by retraining only the last layers of a network. Using this method, a system will be able to perform well despite having relatively few examples. Currently there are several trained models (networks) available and several groups have been successful in starting with trained models such as VGGNet or ResNet, and just training the last layers (fully connected portion) on medical images. This shows the power of CNNs, but we believe that the best results are achieved if the first layers are optimised for medical images. Once such a medical network is available, transfer learning using it would likely retain improved performance. Implementing transfer learning does have some challenges including: (1) the networks are pretrained on images with a specific size, thus forcing the images for the new task to be transformed to match that size; (2) the majority of the pretrained networks are for photographic images, thus three channels are used as input as photographic images use three channels to represent colour (red, green, blue); (3) the range of the input values has to be the same as that used during the training of the original network, i.e., if the image intensities were scales to [−0.5, 0.5] or [0,1].

Data augmentation is a commonly used technique in case of limited or unbalanced datasets both in segmentation and classification tasks. Commonly utilised techniques include: flips, Gaussian noise, jittering, scaling, blurring, rotation, shears, and crops. Generative adversarial network-based techniques have also been proposed,[39] as well as other techniques that leverage the power of CNNs as future generators.[28]

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028

Pitfalls and explainable artificial intelligence

It is important to recognise that the power of deep learning to find signals that predict molecular markers and other important information, but which are not easily perceived can also make it difficult to understand what the algorithms are using to make those predictions. One area of recent activity is focused on understanding the basis for predictions made by artificial intelligence (AI) systems. Early deep learning efforts were criticised because they were viewed as "black boxes" that did not permit users to understand the basis for a given prediction. This has led to a field of inquiry known as "eXplainable AI" or XAI. Understanding the basis for a neural network's prediction has several potential benefits: (1) it may increase confidence in, and acceptance of, the decision/prediction; (2) it may lead to a better understanding of the underlying physical basis (e.g., imaging property that correlates with the genomic property); and (3) there may be legal or regulatory requirements for establishing the basis for decision (e.g., the General Data Protection Regulations recently passed by the European Union). For instance, in the molecular prediction of the biomarkers case, it is important to know if the algorithm decision is based on areas with enhancement or hyperintensities or takes a decision based on the background of the MRI.

In the majority cases, high accuracy is achieved utilising complex models that are not easy to interpret. Additionally artefacts of data or biases can lead to undesirable correlations that the classifiers pick up during training.[40] These issues can be very difficult to identify just by looking at the raw data and predictions. Several methods have been proposed towards interpretation of the predictions. local interpretable model-agnostic explanations (LIME) proposed by Ribeiro *et al.*[40] is a method that interprets individual model predictions based on locally approximating the model around a given prediction. SHAP (SHapley Additive exPlanations) a method proposed by Lundberg *et al.*[41] assigns each feature an importance value for a particular prediction. This method introduces three tools, DeepExplainer (DEEP SHAP), GradientExplainer, and KernelExplainer (Kernel SHAP) for deep learning interpretation. Elenberg *et al.*[42] proposed an approach where the image is first segmented into regions. Subsequently, the network predictions are re-evaluated with most of the image regions replaced by a grey reference image, and the impact on the prediction is recorded. The goal of the approach is to find all the regions that contributed to a decision. Visual explanation such as GradCAM[43] and saliency maps[44] have also been proposed as ways to understand the basis for deep learning network decisions.

It is necessary for deep learning based algorithms to be able to capture both the decision with high accuracy and be able to visualise the areas contributed to this not only to trust the algorithm but also understand the radiological magnification of the biomarkers.

Prognostications

Deep learning has resulted in advances in artificial intelligence technologies that have significantly impacted everyday life, such as self-driving cars and financial investment. The underlying technology is broadly applicable, and that generalisability and the large financial gains have resulted in even greater investments in the basic deep learning technology. These improvements will continue to drive improvements in the performance of medical image systems. In addition, many deep learning fields benefit from understanding the basis for the decision, and medical applications will certainly benefit from those advances, including the improvement of imaging to better demonstrate certain molecular markers, and may help to better understand the biological and pathological basis of those markers in disease.

Deep learning based algorithms are easier to translate in clinical practice thus monotonous tasks, such ROI placement, measurements, and prioritisation would be areas that could be automated saving time and creating better tools for patient management. In addition, although the focus in this paper is the prediction of molecular markers because that is highly valued in clinical practice, it is possible that more comprehensive datasets will enable us to better predict the agent(s) that will be most effective in a given patient, as well as the prognosis for patients with a given imaging appearance. This could increase the value of imaging in the care of patients with brain tumours.

Conflict of interest

The authors declare no conflict of interest.

Funding source

Center of Individualized Medicine, Mayo Clinic.

References

1. Mukherjee S. *A.I. Versus M.D. The New Yorker*. 27 Mar 2017. Available at: https://www.newyorker.com/magazine/2017/04/03/ai-versus-md. [Accessed 13 September 2018].
2. Johnson DR, Diehn FE, Giannini C, *et al*. Genetically defined oligodendroglioma is characterized by indistinct tumour borders at MRI. *AJNR Am J Neuroradiol* 2017;**38**:678–84.
3. Louis DN, Perry A, Reifenberger G, *et al*. The 2016 World Health Organization classification of tumours of the central nervous system: a summary. *Acta Neuropathol* 2016;**131**:803–20.
4. Eckel-Passow JE, Lachance DH, Molinaro AM, *et al*. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumours. *N Engl J Med* 2015;**372**:2499–508.
5. Camara-Quintana JQ, Nitta RT, Li G. Pathology: commonly monitored glioblastoma markers: EFGR, EGFRvIII, PTEN, and MGMT. *Neurosurg Clin N Am* 2012;**23**:237–46 [viii].
6. Eoli M, Menghi F, Bruzzone MG, *et al*. Methylation of O6-methylguanine DNA methyltransferase and loss of heterozygosity on 19q and/or 17p are overlapping features of secondary glioblastomas with prolonged survival. *Clin Cancer Res* 2007;**13**:2606–13.

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028

7. Pallud J, Mandonnet E, Duffau H, *et al*. Prognostic value of initial magnetic resonance imaging growth rates for World Health Organization grade II gliomas. *Ann Neurol* 2006;**60**:380–3.
8. Chamberlain MC. Is the volume of low-grade glioma measurable and is it clinically relevant? *Neuro Oncol* 2014;**16**:1027–8.
9. Gevaert O, Mitchell LA, Achrol AS, *et al*. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology* 2015;**276**:313.
10. Gutman DA, Cooper LAD, Hwang SN, *et al*. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. *Radiology* 2013;**267**:560–9.
11. Zhou H, Vallières M, Bai HX, *et al*. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol* 2017;**19**:862–70.
12. Zhang X, Tian Q, Wang L, *et al*. Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI. *J Magn Reson Imaging* 2018;**48**(4):916–26. https://doi.org/10.1002/jmri.25960.
13. Jakola AS, Zhang Y-H, Skjulsvik AJ, *et al*. Quantitative texture analysis in the prediction of IDH status in low-grade gliomas. *Clin Neurol Neurosurg* 2018;**164**:114–20.
14. Li Z-C, Bai H, Sun Q, *et al*. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur Radiol* 2018;**28**:3640–50.
15. Park YW, Han K, Ahn SS, *et al*. Whole-tumour histogram and texture analyses of DTI for evaluation of IDH1-mutation and 1p/19q-codeletion status in World Health Organization grade ii gliomas. *AJNR Am J Neuroradiol* 2018;**39**:693–8.
16. Chang K, Bai HX, Zhou H, *et al*. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. *Clin Cancer Res* 2018;**24**:1073–81.
17. Rathore S, Akbari H, Rozycki M, *et al*. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci Rep* 2018;**8**:5087.
18. van der Voort SR, Gahrmann R, van den Bent MJ, *et al*. Radiogenomic classification of the 1p/19q status in presumed low-grade gliomas. In: *2017 IEEE 14th international symposium on biomedical imaging*. ISBI 2017; 2017, https://doi.org/10.1109/isbi.2017.7950601.
19. Han L, Kamdar MR. MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. *Pac Symp Biocomput* 2018;**23**:331–42.
20. Korfiatis P, Kline TL, Lachance DH, *et al*. Residual deep convolutional neural network predicts MGMT methylation status. *J Digit Imaging* 2017;**30**:622–8.
21. Korfiatis P, Kline TL, Erickson BJ. Evaluation of a deep learning architecture for MR imaging prediction of ATRX in glioma patients. In: *SPIE medical imaging, 2018: computer-aided diagnosis*; 27 February. 2018, https://doi.org/10.1117/12.2293538. Houston, Texas.
22. Chang P, Grinband J, Weinberg BD, *et al*. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am J Neuroradiol* 2018;**39**(7):1201–7. https://doi.org/10.3174/ajnr.A5667.
23. Eichinger P, Alberts E, Delbridge C, *et al*. Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas. *Sci Rep* 2017;**7**:13396. https://doi.org/10.1038/s41598-017-13679-4.
24. Bisdas S, Shen H, Thust S, *et al*. Texture analysis- and support vector machine-assisted diffusional kurtosis imaging may allow in vivo gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci Rep* 2018;**8**:6108.
25. Ren Y, Zhang X, Rui W, *et al*. Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features. *J Magn Reson Imaging* 2018, https://doi.org/10.1002/jmri.26240.
26. Pan C-C, Liu J, Tang J, *et al*. A machine learning-based prediction model of H3K27M mutations in brainstem gliomas using conventional MRI and clinical features. *Radiother Oncol* 2018, https://doi.org/10.1016/j.radonc.2018.07.011.
27. Arita H, Kinoshita M, Kawaguchi A, *et al*. Lesion location implemented magnetic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. *Sci Rep* 2018;**8**:11773.
28. Liu J, Chen F, Pan C, *et al*. A Cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas. *IEEE Trans Biomed Eng* 2018;**65**:1943–52.
29. AMIA. *Informatics summit - session details*. 2018. Available at: https://informaticssummit2018.zerista.com/event/member?item_id=7574587. [Accessed 17 September 2018].
30. Fetit AE, Novak J, Peet AC, *et al*. Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours. *NMR Biomed* 2015;**28**:1174–84.
31. Fetit AE, Novak J, Rodriguez D, *et al*. Radiomics in paediatric neuro-oncology: a multicentre study on MRI texture analysis. *NMR Biomed* 2018;**31**, https://doi.org/10.1002/nbm.3781.
32. Perreault S, Ramaswamy V, Achrol AS, *et al*. MRI surrogates for molecular subgroups of medulloblastoma. *AJNR Am J Neuroradiol* 2014;**35**:1263–9.
33. Keil VC, Warmuth-Metz M, Reh C, *et al*. Imaging biomarkers for adult medulloblastomas: genetic entities may be identified by their MR imaging radiophenotype. *AJNR Am J Neuroradiol* 2017;**38**:1892–8.
34. Coroller TP, Bi WL, Huynh E, *et al*. Radiographic prediction of meningioma grade by semantic and radiomic features. *PLoS One* 2017;**12**:e0187908.
35. Phuttharak W, Boonrod A, Thammaroj J, *et al*. Preoperative MRI evaluation of meningioma consistency: a focus on detailed architectures. *Clin Neurol Neurosurg* 2018;**169**:178–84.
36. Suh HB, Choi YS, Bae S, *et al*. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. *Eur Radiol* 2018;**28**:3832–9.
37. Zacharaki EI, Wang S, Chawla S, *et al*. Classification of brain tumour type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med* 2009;**62**:1609–18.
38. Citak-Er F, Firat Z, Kovanlikaya I, *et al*. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. *Comput Biol Med* 2018;**99**:154–60.
39. Antoniou A, Storkey A, Edwards H. *Data augmentation generative adversarial networks. arXiv [stat.ML]*. 2017. Available at: http://arxiv.org/abs/1711.04340.
40. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM; 2016. p. 1135–44.
41. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, *et al*., editors. *Advances in neural information processing systems 30, 4-9 december 2017, long beach, California*. NIPS/Curran Associates; 2017. p. 4765–74.
42. Elenberg E, Dimakis AG, Feldman M, *et al*. Streaming weak submodularity: interpreting neural networks on the fly. In: Guyon I, Luxburg UV, Bengio S, *et al*., editors. *Advances in neural information processing systems 30, 4-9 december 2017, long beach, California*. NIPS/Curran Associates; 2017. p. 4044–54.
43. Chattopadhyay A, Sarkar A, Howlader P, *et al*. *Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. arXiv [cs.CV]*. 2018. Available at: http://arxiv.org/abs/1710.11063.
44. Simonyan K, Vedaldi A, Zisserman A. *Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv [cs.CV]*. 2013. Available at: http://arxiv.org/abs/1312.6034.

Please cite this article as: Korfiatis P, Erickson B, Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.01.028